A Research Project Approved and Financed by SUST Research Center, Research Grant 2018-19, (Project Code: PS/2018/1/13), Shahjalal University of Science and Technology, Sylhet – 3114

# PROJECT TITLE

An Evaluation of Time Series Models and Machine Learning Techniques for Predicting Global Warming: Evidence from Bangladesh.

|  |  |
| --- | --- |
| DR. KHALIDUR RAHMAN  (Principal Investigator)  Professor  Department of Statistics  Shahjalal University of Science and Technology  Sylhet – 3114 |  |



SHAHJALAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

SYLHET – 3114, BANGLADES

A Research Project Approved and Financed by SUST Research Center, Research Grant 2018-19, (Project Code: PS/2018/1/13), Shahjalal University of Science and Technology, Sylhet – 3114

PROJECT TITLE

An Evaluation of Time Series Models and Machine Learning Techniques for Predicting Global Warming: Evidence from Bangladesh.

SUBMITTED BY

|  |  |  |  |
| --- | --- | --- | --- |
| DR. KHALIDUR RAHMAN  (Principal Investigator) |  |  |  |
| Professor  Department of Statistics  Shahjalal University of Science and Technology  Sylhet – 3114 | |  |  |

ASISSTANCE BY

MOHAMMAD NAYEEM HASAN

(Research Assistant)

MS Student

Registration No. 2016124006

Session: 2016-17

Department of Statistics

Shahjalal University of Science and Technology, Sylhet – 3114

SHAHJALAL UNIVERSITY OF SCIENCE AND TECHNOLOGYSYLHET – 3114, BANGLADESH

# DECLARATION

This project is a record of an original research work. We proclaim that all sources used in exploring it are fully acknowledged and all quotations properly identified. The work has been done according to the instruction and guidance of Research Center, Shahjalal University of Science and Technology, Sylhet.

We hereby also declare that this project report or any part of it has not been submitted elsewhere to obtain any financial assistance or project grant or published any time before. We understand the ethical implications of our research, and this work meets the requirements of the Research Center, Shahjalal University of Science and Technology, Sylhet.

|  |  |  |
| --- | --- | --- |
| DR. KHALIDUR RAHMAN  (Principal Investigator) |  |  |
| Professor  Department of Statistics  Shahjalal University of Science and Technology  Sylhet – 3114 | |

# ACKNOWLEDGEMENT

All praises are due to Almighty Allah, who enabled me to complete the present piece of project work successfully and to submit the project at the SUST Research Center.

I would like to thank SUST Research Center for giving such attention and providing fund for this work. I am particularly grateful to Professor Dr. Zakir Hossain (Director, SUST Research Center) for his appreciation and suggestion (typically to follow the financial rules) during the project period.

It is indeed my privilege to express my deep sense of gratitude and indebtedness to my co-investigator Tajmunnaher, Assistant Professor, Department of Civil & Environmental Engineering, Shahjalal University of Science and Technology, Sylhet for her immense help and co-operation throughout the tenure of the project work.

I would profoundly like to appreciate Mohammad Nayeem Hasan, a MS student of the Department of Statistics, Shahjalal University of Science and Technology, Sylhet for his valuable works and continuous assistance in this project.

Finally, I wish to acknowledge my beloved family members for their continuous support and inspiration for research works.

Thank You,

|  |
| --- |
| Dr. Khalidur Rahman  (Principal Investigator) |
| Professor, Department of Statistics  Shahjalal University of Science and Technology, Sylhet – 3114 | |

# ABSTRACT

Climate change has become a significant worldwide problem since it affects all animals and plants on the earth. Rising temperatures can lead to more frequent and severe heatwaves, droughts, and storms. Bangladesh is regarded as one of the most exposed nations facing the difficulties of climate change and the effects of global warming. Predicting temperature has been a critical climatic need for several applications in industries including agriculture, manufacturing, energy, environment, tourism, etc. The purpose of this study was to assess the capacity of several machine learning (ML) techniques to forecast temperature to investigate climate change, as well as to compare them to classical time series models. To forecast and predict temperature, time series models such as Autoregressive Integrated Moving Average (ARIMA), Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX), and Machine learning techniques, such as k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forests (RF), Adaptive Boosting (AdaBoost or AdB), Artificial Neural Network (ANN) are used in this study. Data were collected between January 1, 1965, and December 31, 2021, or approximately 57 years for Sylhet. Trends in the data were examined using the Mann-Kendall test. Except for ARIMA and SVM, all the methods were well-fitted, according to the test findings. The following Seven performance indicators were used to determine the best-fitted method: R2, RSR, NSE, MAE, Index of Agreement (d), PBIAS (%), and RMSE. However, these indicators showed that ANN was the best method compared to others for predicting temperature. More research on temperature prediction for large regions using a similar mechanism is required.

Major limitation: due to the budget constraint, only the Sylhet region consider for our study.

Table of Contents

[PROJECT TITLE i](#_Toc123581641)

[DECLARATION ii](#_Toc123581642)

[ACKNOWLEDGEMENT iii](#_Toc123581643)

[ABSTRACT iv](#_Toc123581644)

[CHAPTER I 7](#_Toc123581645)

[1 Introduction 7](#_Toc123581646)

[1.1 Background 7](#_Toc123581647)

[1.2 Problem Statement 9](#_Toc123581648)

[1.3 Objectives of the Study 10](#_Toc123581649)

[1.4 Rationale of the Study 10](#_Toc123581650)

[1.5 Limitations of Study 10](#_Toc123581651)

[CHAPTER II 11](#_Toc123581652)

[2 LITERATURE REVIEW 11](#_Toc123581653)

[2.1 Preamble 11](#_Toc123581654)

[2.2 Previous Studies 11](#_Toc123581655)

[CHAPTER III 13](#_Toc123581656)

[3 METHODOLOGY 13](#_Toc123581657)

[3.1 Study Area 13](#_Toc123581658)

[3.2 Data Collection and Preprocessing 13](#_Toc123581659)

[3.3 Statistical Methods 13](#_Toc123581660)

[3.4 Machine-Learing Methods 15](#_Toc123581661)

[CHAPTER IV 19](#_Toc123581662)

[4 Result and Discussion 19](#_Toc123581663)

[4.1 Time Series Models Diagnostics 19](#_Toc123581664)

[4.2 Machine Learning Models Diagnostics 22](#_Toc123581665)

[4.3 Comparison between time series models and Machine Learning Models 23](#_Toc123581666)

[CHAPTER V 31](#_Toc123581667)

[5 CONCLUSIONS AND RECOMMENDATIONS 31](#_Toc123581668)

[5.1 Introduction 31](#_Toc123581669)

[5.2 Findings of the Research 31](#_Toc123581670)

[5.2.1 Time Series Models 31](#_Toc123581671)

[5.2.2 Machine learning Methods 31](#_Toc123581672)

[5.3 Comparison between Statistical Methods and Machine Learning Methods 32](#_Toc123581673)

[5.4 Limitations and Recommendations for Future Works 32](#_Toc123581674)

[5.4.1 Limitations of the Study 32](#_Toc123581675)

[5.4.2 Recommendations 32](#_Toc123581676)

[REFERENCES 33](#_Toc123581677)

|  |  |  |
| --- | --- | --- |
| LIST OF TABLES | | |
| Table No. | Table Name | Page |
| Table 4.1a | Augmented Dickey-Fuller Test after differencing. | 20 |
| Table 4.1b | Estimated parameters of ARIMA (1, 2, 2) model. | 21 |
| Table 4.1c | Estimated parameters of ARIMA (1, 2, 2) model with exogenous variables. | 21 |
| Table 4.2a | Features collinearity and correlation matrix. | 12 |
| Table 4.2b | Mann Kendall test statistics | 13 |
| Table 4.3 | Performance ratings of the models for the test dataset. | 22 |

|  |  |  |  |
| --- | --- | --- | --- |
| LIST OF FIGURES | | | |
| Figure No. | Figure Name | Page | |
| Figure 3.4 | Presentation of a basic artificial neuron | | 18 |
| Figure 4.1a | The trend of Temperature over time | | 19 |
| Figure 4.1b | Autocorrelation plot | | 20 |
| Figure 4.1c | Autocorrelation and Partial autocorrelation plot after differencing. | | 20 |
| Figure 4.3a | Relationship between observed and test outputs for ARIMA | | 24 |
| Figure 4.3b | Relationship between observed and test outputs for ARIMAX | | 25 |
| Figure 4.3c | Relationship between observed and test outputs for KNN | | 26 |
| Figure 4.3d | Relationship between observed and test outputs for SVM | | 26 |
| Figure 4.3e | Relationship between observed and test outputs for Random Forest | | 27 |
| Figure 4.3f | Relationship between observed and test outputs for AdB | | 28 |
| Figure 4.3g | Relationship between observed and test outputs for ANN | | 29 |

# CHAPTER I

## 1 Introduction

## 1.1 Background

Climate change has become a major global concern since it is affecting substantially all the plants and animals on the planet (Tol, 2013). Mitigating climate change is one of the biggest challenges of humankind. Despite the complexity of predicting the effects of climate change on earth, there is a scientific consensus about its negative impacts. Among them, the affectation of ecosystems, decrease of biodiversity, soil erosion, extreme changes in temperature, sea level rise, and global warming have been identified. Likewise, impacts on the economy, human health, food security, and energy consumption are expected (Abdel-Aal, 2004; Pachauri et al., 2014). Specifically, temperature forecasting has been a crucial climatic factor required for many different applications in areas such as agriculture, industry, energy, environment, tourism, etc (Li et al., 2016). Weather forecasting is supposed to be a prime factor for Bangladesh’s economy as agriculture plays a vital role in the country’s overall Gross Domestic Product (GDP) which accounts for approximately 20% of the total amount. Nearly 70% of its total population live in rural areas and 60% of them earn their livelihood from tillage stuff. Even a significant amount of total annual exports are from farming products which tend to be in the region of 13-18% of the country’s total GDP. So, the discrepancy of rainfall, humidity, wind speed, the temperature in time, space, and aggregation affects the country’s agriculture which might hamper the economy to a greater extent (Islam & Uyeda, 2007; Khan et al., 2015; Rahman et al., 2017; Rina et al., 2012). Therefore, there is a need to accurately predict temperature values because, in combination with the analysis of additional features in the subject of interest, they would help to establish a planning horizon for infrastructure upgrades, insurance, energy policy, and business development (Smith et al., 2007).

The most influential climate elements were evaporation, minimum humidity, and sunshine duration(Sawan, 2018). Due to high rates of evaporation, Australia loses over 40% of its entire water storage capacity each year (Helfer et al., 2012). The most often used humidity measurement has historically been dew point temperature. It is a crucial instrument for predicting both relative and specific humidity. Therefore, a relative humidity change follows a dew point temperature change(Mortuza et al., 2014). Climate scientists correlate the Clausius-Clapeyron equation to anthropogenic global warming and demonstrate that relative humidity boosts the atmosphere's ability to store water by around 7% for every 1 degree of warming (Schaller et al., 2016). A worldwide rising trend of cloud cover, near-surface humidity, and precipitation has been correlated with higher temperatures and enhanced evapotranspiration(Huntington, 2006).  Cloud cover, specific humidity, and precipitation increased in areas where nighttime temperatures rose by >0.5°C above daylight values. In contrast, cloud cover, specific humidity, and precipitation decreased where daylight temperatures rose by >0.5°C more than nighttime temperatures(Cox et al., 2020). The decade 1991-2000 had an increase in precipitation (6%) and mean annual warming (0.8°C) over Sweden compared to the years 1961-1990. According to model simulations, there is a likelihood of 6-7 percent that these changes only happened due to natural variability and that the observed discrepancies were caused by a mix of anthropogenic forcing and natural climate variability (Räisänen & Alexandersson, 2003). In low elevations, the concurrent impact of changes in atmospheric circulation and a rise in its moisture content will cause a simultaneous reduction in sea level of between 2 and 4 cm. For quadruple atmospheric CO2 concentrations, corresponding rates of sea level rise in subpolar areas range from 0.4 to 0.6 mm per year, whereas in mid- and low latitudes, sea level will fall at a rate of 0.2 mm per year (Stammer & Hüttemann, 2008).

Implementations in meteorology and other environmental fields use time series analysis and forecasting as key tools to comprehend characteristics like rainfall, humidity, temperature, draught, and other associated consequences (Nury et al., 2013). To forecast and predict temperature, time series models such as Auto Regressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) are mostly used (Bang et al., 2019; Colak et al., 2015; Dimri et al., 2020; Shakti et al., 2017). Machine learning techniques, such as support vector machines (SVM), artificial neural networks (ANN), multivariate discriminant analyses (MDA), etc., are effective modeling tools for establishing a connection between inputs and outputs because they are not constrained by the same presumptions as statistical techniques (such as ARMA and ARIMA) (Chou & Tran, 2018). Using a collection of input characteristics, including historical data for temperature, relative humidity, solar radiation, rain, and wind speed measurements, machine learning algorithms can improve to forecast temperatures correctly, among others(Cifuentes et al., 2020). Especially when addressing climate change, machine learning (ML) algorithms for time series prediction are improving in accuracy and value (Azari et al., 2022). Linear Regression is one of the established and well-known ML techniques used in various environmental fields. It may be used to solve issues with one or more variables (Ahmadi et al., 2022; Marill, 2004; Mehri et al., 2021). K-nearest Neighbor (kNN), another nonparametric ML technique, has a long history of application in ML literature to address regression issues [38-40]. Regression using the Support Vector Machine (SVM) is another technique used to identify relationships between inputs and outputs (Liu et al., 2013; Wang et al., 2010). The Artificial Neural Network (ANN), which can recognize nonlinear patterns in functional connection characteristics and targets, is one of the most well-known and widely used techniques. It has been used to solve several ML issues in various fields (Fotovvati et al., 2020; Najah et al., 2013; Singh et al., 2009). Six machine learning techniques were examined in Iranian research, including linear regression, k-nearest neighbor, support vector machines, artificial neural networks, random forests, and adaptive boosting. The method's overall efficacy in predicting the average daily temperature value and other outcomes was assessed by comparing all the training and predictive performances. The results showed that the Artificial Neural Network outscored the other five evaluated techniques in both training and temperature value prediction (Azari et al., 2022). Neural network models can be used as promising techniques to forecast air temperature, according to a review of Neural Networks for Air Temperature Forecasting. Although ANN-based algorithms have been widely used to estimate air temperature because of their quick computation speed and capacity to handle complicated issues, there is still no agreement on the most effective technique currently in use (Tran et al., 2021).

Bangladesh is regarded as one of the most exposed nations facing the difficulties of climate change and the effects of global warming due to its geographic position, low elevation of sea level, high population density, overreliance on natural resources, and insufficient ability to cope with the concerns (Al-Amin et al., 2013). About one-third of the nation would be submerged by a one-meter sea level rise, forcing 25–30 million Bangladeshis to leave their homes(Choudhury et al., 1997). Based on projections from several climate models and an assessment of long-term records of temperature and rainfall at a few locations in Bangladesh, the problem of global warming and sea level rise related to the greenhouse effect is addressed. A comparison of the IPCC report and certain model findings shows that sea level rise and other climatic changes would likely occur in this region in the future because of greenhouse warming. The socio-economic status of Bangladesh is directly harmed by changes in climatic variables like temperature and rainfall as well as related issues like increased surface warming, floods, sea level rise, etc (Choudhury et al., 1997).

## 1.2 Problem Statement

The world average temperature is expected to rise by 1.5°C and 2.5°C, respectively, by the middle and end of the twenty-first century. By 2030, 2050, and 2100, respectively, these anticipated temperatures will cause sea levels to increase by about 14, 32, and 88 cm, inundating 8, 10, and 16 percent of Bangladesh's entire landmass, respectively. Most of the coastal regions and related islands in the Khulna, Barisal, and western Chattagram divisions are located within one meter of sea level, where saltwater incursion is frequent. It is anticipated that the projected sea level rise would cause these regions to be swamped and unsuitable for crop cultivation (Awal & Khan, 2020). Therefore, accurate modeling of the local temperature based on climate factors at the global level should be done to order to mitigate the negative effects of global warming on Bangladesh. Although a few research (Pour et al., 2018; Shafin, 2019; Zaman, 2018) have documented the application of machine learning techniques on meteorological variables, no study has used machine learning models that we consider in our study to investigate global warming in Bangladesh.

## 1.3 Objectives of the Study

The present study, filling in the aforementioned research gap, is mainly

aimed to investigate global warming (temperature) in Bangladesh for almost 57 years

based on daily data collected from different public sources. The specific objectives are:

1. To employ the classical time series model ARIMA, ARIMAX in the temperature data.
2. To employ and compare machine learning approaches to explore the effects of different influences on global warming in Bangladesh.
3. To forecast the temperature as well as the concentration of several climate variables for the next several years.
4. To visualize the real, predicted, and forecasted data so that the actuality could be understood briefly.

## 1.4 Rationale of the Study

Over the past ten years, there have been more global and regional efforts to comprehend the impact of historical climate change. Studies on the effects of climate change on the agricultural, ecological, environmental, and industrial sectors have focused particularly on estimates of air temperatures as a critical element. A precise temperature forecast is essential for planning actions for the government, business, and public to protect people and property. As more methods are evolved for time series prediction, it also becomes more crucial differentiating between them. This research will help policymakers to take the necessary steps to face challenges that are being shaped by climate change. Thus, it is directly linked with SDG#13 Climate Action.

## 1.5 Limitations of Study

Although this study was carefully prepared, the researchers were aware of its limitations and drawbacks.

* Only the Sylhet region was taken into consideration for our study because of budgetary constraints.
* Reanalysis data was used because there was a lack of sufficient data in the Bangladesh Meteorological Department(BMD).

# CHAPTER II

## 2 LITERATURE REVIEW

### 2.1 Preamble

This chapter's major goal is to provide a conceptual framework summary for the evaluation of time series models and machine learning methods for global warming prediction.

### 2.2 Previous Studies

One of the major challenges confronting humans is reducing the effects of climate change. Despite the difficulty in foreseeing climate change's consequences on the planet, scientists agree that it will have detrimental implications. Ecosystem disruption, a decline in biodiversity, soil erosion, abrupt temperature fluctuations, an increase in sea level, and global warming have been named as some of these. Similar effects are anticipated for the economy, human health, food security, and energy use (Pachauri & Reisinger, 2008; Tol, 2002). Predicting air temperature has been a critical climatic need for several applications in industries including agriculture, manufacturing, energy, environment, tourism, etc. (Abdel-Aal, 2004). In many applications in meteorology and other environmental fields, time series analysis and forecasting have emerged as key tools for comprehending phenomena like rainfall, humidity, temperature, draught, etc.

In the Sylhet and Moulvibazar districts of northeast Bangladesh, ARIMA (Auto Regressive Integrated Moving Average) models have been put up and utilized to carry out short-term forecasts of monthly maximum and minimum temperatures. Set up stationary, seasonal ARIMA models for the temperatures measured at two sites in the Sylhet division between 1977 and 2011 using the traditional Box-Jenkins approach. For the years 2010 to 2012, the models have been verified. The best orders of the ARIMA models are chosen and assessed using the AIC criterion based on the examination of the ACF and PACF autocorrelation plots (Nury et al., 2013). Three forecasting techniques are utilized in another study: ARMA (Auto Regressive Moving Average), SARIMA (Seasonal Auto-Regressive Integrated Moving Average), and ARMAX (ARMA with Exogenous Variables). Based on a comparison of the three models' performances, the best model is used to forecast temperature and rainfall, which are then utilized to forecast crop production using a fuzzy logic model (Bang et al., 2019). Another study revealed a model that, when using the Seasonal ARIMA, TBATS, Seasonal Reg-ARIMA, and NNETAR models, can predict the daily peak load for the next day when taking temperature and weekday, weekend, and holiday influences into account. The findings of this paper's model's forecasting performance test for a Seasonal Reg-ARIMA model and NNETAR model that can take the day of the week and temperature into account revealed greater forecasting performance than a model that cannot take these aspects into account (Lee & Kim, 2019).

A review of temperature forecasting using machine learning techniques demonstrates that these techniques can aid in the accurate prediction of temperatures based on a set of input features, which can include, among other things, historical readings for temperature, relative humidity, solar radiation, rain, and wind speed. In comparison to conventional Artificial Neural Network (ANN) designs, Deep Learning techniques exhibit fewer errors (Mean Square Error = 0.0017 °K) for 1 step ahead at a regional scale. Support Vector Machines (SVM) are favored on a worldwide scale because of their excellent balance between simplicity and precision(Cifuentes et al., 2020). Another study employed three machine learning models for temperature prediction through a comparative analysis utilizing meteorological data gathered from Central Kerala between 2007 and 2015: Multiple Linear Regression (MLR), Artificial Neural Network (ANN), and Support Vector Machine (SVM). Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Correlation Coefficients are used to assess the experimental outcomes (CC). The CC and error metrics demonstrate that MLR is a more accurate model for predicting temperature than ANN and SVM (Anjali et al., 2019).

A comparative study employed a set of pertinent climatology time series as independent variables and used temperature time series as a dependent variable. Four basic models—Linear Regression (L.R.), k-Nearest Neighbor (kNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN)—as well as two ensemble model methods—Random Forest (R.F.) and Adaptive Boosting—were among the six techniques evaluated (AdB). The method's overall efficacy in predicting the average daily temperature value was assessed by comparing all of the training and prediction performances. Each of the ML techniques discussed was trained on real data before being applied to forecast the temperature in the research region. The analysis showed that the Artificial Neural Network beat the other five evaluated techniques in both training and forecasting temperature (Azari et al., 2022).

# CHAPTER III

## 3 METHODOLOGY

### 3.1 Study Area

The administrative division of Sylhet in north-eastern Bangladesh is situated at 24.8917°N latitude and 91.8833°E longitude. The Sylhet District and its surroundings are covered by this station. Due to its abundant natural resources, like its tilas (little hills), which contain more than 150 tea plantations, the northeastern region of Bangladesh makes for an intriguing research area. An increasing population that depends on the wetland for their livelihood, as well as the general well-being of a variety of animals, is supported in the area.

### 3.2 Data Collection and Preprocessing

The collected raw dataset from BMD and ERA 5 hourly data was preprocessed and a lot of cleaning was required to convert it from a semi-structured dataset to a structured dataset and prepared for the implementation of our desired model. After completing the preprocessing of the dataset to make the forecasting where each dataset had 4 specific parameters e.g., wind speed, rainfall, humidity, and temperature (low and high). We conducted our research on this preprocessed dataset and later compared all the outcomes for better predictions. Provided data had the category of rainfall (millimeters), humidity (percentage of water in the air), wind speed (kilometer per hour), and temperature (degree Celsius). Basically, long-range forecasting can be divided into 4 categories, (a) periodicity approach (b) correlation approach (c) extended synoptic approach, and (d) dynamical approach (Rajeevan et al., 2007). BMD is a government agency for weather prediction in Bangladesh. In 2007 BMD first introduced a statistical forecast system based on an ensemble technique (Andrade & Bessa, 2017; Mahabub et al., 2019). Although their predictions were acceptable that was always dependent on some specific predictors. However, we tried to put it one step forward through our research.

### 3.3 Statistical Methods

#### 3.3.1 ARIMA

An [ARIMA model](https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average) is a class of statistical models for analyzing and forecasting time series data. It explicitly caters to a suite of standard structures in time series data, and as such provides a simple yet powerful method for making skillful time series forecasts (Time Series Analysis, n.d.). Since then, ARIMA models have been widely applied in a wide range of time series analysis applications. According to research done by Chen & Lai, 2011 which involves wind speed forecasting by applying the ARIMA models methodology of box and Jenkins as follows,

Step 1: Model Recognition

Step 2: Estimation of parameters

Step 3: Checking the residual diagnostics and forecasting

Model Recognition:

ARIMA models are one of the most widely used approaches for time series forecasting. For a time series that is stationary ARIMA (p, d, q) Model can be written in terms of past temperature data, residuals, and prediction errors as follows

x= Temperature time series data

x(t-i) = (t-i)th data

a(t)= random white noise time series

= auto-regressive parameters

=moving average parameters

p=order of auto-regressive model

q=order of moving average model

d=degree of differencing

Estimation of parameters

Parameter estimation involves choosing the right p and q which can describe the time series with the highest accuracy. These parameters can be approximately estimated by analyzing the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The ACF function can be helpful to determine the order of moving average q and the PACF function can be used to determine the order of the auto-regressive model. P value is the lag value where the PACF graph crosses the upper confidence gap. The Q value is the lag value where the ACF graph crosses the upper confidence gap.

Checking the residual diagnostics:

Residuals are leftovers after fitting an appropriate model. In most cases, residuals can be regarded as the difference between the actual values and the forecasted values.

A residual diagnostic is performed to check whether the model has been able to capture adequate information in the time series data. For appropriate forecasting, model residuals should be uncorrelated and should have zero mean.

#### 3.3.2 ARIMAX

An Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) model can be viewed as a multiple regression model with one or more autoregressive (AR) terms and/or one or more moving average (MA) terms. This method is suitable for forecasting when data is stationary/nonstationary and multivariate with any type of data pattern. ARIMAX is related to the ARIMA technique, while ARIMA is suitable for univariate datasets. ARIMAX is suitable for analysis where there are additional explanatory variables (multivariate) in categorical and/or numeric format.

### 3.4 Machine-Learing Methods

#### 3.4.1 KNN

k-Nearest Neighbors or KNN is a non-parametric method that is used for both classification and regression-type problems. In the case of KNN regression, the output is the property value for the object. This value is the average of the values of k-Nearest Neighbors. It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. It is supposed to be the simplest of all ML algorithms. The KNN algorithm is used for estimating continuous variables. It uses a weighted average of the k nearest neighbors, weighted by the inverse of their distance. First of all, the Euclidian or Mahalanobis distance is computed from the query example to the labeled examples. Then the labeled examples are ordered by increasing distance. In the next step, it is tried to find a heuristically optimal number of k of nearest neighbors, based on the values of Root Mean Squared Error (RMSE). This is done using cross-validation. In the end, an inverse distance weighted average is calculated with the k-nearest multivariate neighbors (Akter et al., 2009).

#### 3.4.2 SVM

Support Vector Regression (SVR) is used to predict the maximum temperature at a location. Support vector regression is different from conventional regression techniques because it uses Structural Risk Minimization (SRM) but not the Empirical Risk Minimization (ERM) induction principle which is equivalent to minimizing an upper bound on the generalization error and not the training error. Due to this feature, it is expected to perform better than conventional techniques which may suffer from possible overfitting. In this work Maximum Temperature of a day is predicted based on the maximum temperature of the previous n days where n is the optimal length of the span. The value of n is found by experimentation. The available data is divided into training, validation, and test sets. The training set is used to build the model, the validation set is used for parameter optimization and the test set is used to evaluate the model. A nonlinear support vector regression method is used to train the SVM. Support Vector Regression or SVR is based on statistical learning theory. It is widely used in both classification and regression types of problems. Its most efficiency is observed in forecasting financial data and time series prediction. The SVR must use a cost function to measure the estimated risk to lessen the regression error. One might choose a loss function to calculate the cost from the least module loss function, quadratic loss function, etc. The insensitive loss function exhibits the sparsity of the solution. It contains a fixed and symmetrical margin term. It runs into the risk of overfitting the data with poor generalization if the margin is either zero or very small. On the contrary, if the margin tends to be large, it gains a better generalization at the risk of having a higher testing error. Generally, the estimation function in SVR takes the following form (Yang et al., 2002):

In this Eq. (.) denotes the inner product in Ω, a feature space of possibly different dimensionality such that ω: X → Ω and b ∈ R. The other two parameters, ω, and b can be determined from the training dataset by minimizing the regression risk based on the estimated risk.

#### 3.4.3 Random Forest

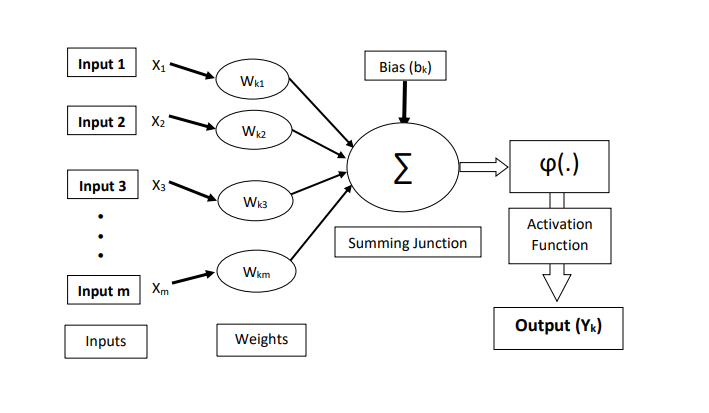
Random Forests (RF) are one of the best-known bagging algorithms that are suggested by Breiman, 2001. Rfs are tree-based EML algorithms that use decision trees as base learners so that each tree is dependent on a set of random parameters. Each tree is created from a bootstrap sample of the primary dataset. RFs can be employed for either a categorical dependent parameter (response variable) for classification purposes or a continuous response variable, for developing regressive models. RFs, similar to other EML algorithms, combine multiple individual models to make predictions. Therefore, this method can be a better choice than a single decision tree since it decreases the over-fitting by averaging the result. The RF algorithm process has several stages: Firstly, bootstrap samples are randomly selected from the given original dataset. Then a decision tree is created for each sample, and the prediction result of each decision tree is obtained in this stage. Next, for each predicted result, the voting step is carried out. Eventually, the most popular prediction model is chosen as the final prediction result (Breiman, 2001).

Random Forests (R.F.s) are bagging ensemble models that use several decision trees as base learners to produce more accurate results. Individual trees are created by the bootstrap sampling process from training data and use a set of random parameters as their roots and nodes. Multiple decision trees provide stability over a single tree by reducing over-fitting and averaging the results(Ardabili et al., 2020). The three main parameters for the R.F.s are the number of trees in the forest at each binary node, the number of randomly selected predictors, and the minimal number of observations at the nodes of the trees (Pham et al., 2021).

#### 3.4.4 AdB

Boosting is a generic and demonstrably effective technique for combining imprecise and comparatively inaccurate rules of thumb to produce a very accurate prediction rule. AdB is a boosting technique that can use multiple RT learners as its core learners. The AdB begins the procedure by producing core learners (RT) with lower precision and then enhances the successive learners based on the outcomes of the preceding learners. The algorithm creates weights for the data points during this procedure. AdB begins with a uniform weight distribution across all of the subsample's data points. As the process goes on, various weights are given to every data item according to how crucial they are to the model's accuracy. The original training dataset is used to test the RT learners trained using bootstrapped datasets, revealing data points that were incorrectly predicted. In order to increase the accuracy of the subsequent learner, the AdB gives preference to those data points that were incorrectly predicted (Duan et al., 2019; Mousavi et al., 2019).

#### 3.4.5 ANN (RELU)

The Artificial Neural Network (ANN) is a popular machine learning technique that has been used successfully to solve challenging problems in the areas of prediction, pattern recognition, simulation, optimization, and several others (Mohamed, 2019). The ANN implements the error-backpropagation learning technique to capture the nonlinear pattern between the input and output (Hastie et al., 2001). An ANN network is a collection of input and output layers, as well as a hidden layer with connections.

**Figure-3.4 : Presentation of a basic artificial neuron**

A neuron, which serves as the foundational processing component of a NN, collects inputs and generates output. As indicated in **Figure 3.4**, each input is multiplied by connection weights, the products and biases are added, and the output is moved through an activation function(Mohamed, 2019). The trial-and-error methodology should be used to determine the total number of hidden layers and the number of neurons in each hidden layer (Rasamoelina et al., 2020). In the present studies, for the hidden layers, the RELU activation function was employed. If x is less than 0 in the RELU activation function, the output is set to 0, and if it is more than 0, it becomes x, which determines the output of the neural network, such as yes or no. The resultant values are mapped by the activation function between 0 and 1, -1 to 1, etc. (depending upon the function). Then again, the output layer receives the output from the hidden layers (Rasamoelina et al., 2020).

# CHAPTER IV

## 4 Result and Discussion

### 4.1 Time Series Models Diagnostics

4.1.1 ARIMA:

ARIMA Modeling process starts with observing the data over time. In our case, visualizing the temperature against time gives us a clear indication of the trend. In the next step, stationarity

Chart, bar chart

Description automatically generated

**Figure 4.1a: Trend of Temperature over time**

of data needs to be checked. The autocorrelation plot shows a slow decay which means the data is clearly not stationary. We perform the differencing 2 times to make the data stationary. Data becomes stationary as we can see after the differencing this autocorrelation approach zero

Chart

Description automatically generated

**Figure 4.1b: Autocorrelation plot**

exponentially after the first or second lag. The formal Augmented Dickey-Fuller test also shows that the data become stationary after taking the differencing with a p-value of 0.01. Now we can

Chart, scatter chart

Description automatically generatedChart

Description automatically generated

**Figure 4.1c: Autocorrelation and Partial autocorrelation plot after differencing.**

|  |  |
| --- | --- |
| Augmented Dickey-Fuller Test | |
| Alternative Hypothesis: Data is stationary | |
| Dickey-Fuller | -54.818 |
| Lag order | 27 |
| p-value | 0.01 |

**Table 4.1a: Augmented Dickey-Fuller Test after differencing.**

estimate the ARIMA model. This estimation includes a degree of autoregressive (p), differencing (d), and moving average (q). We already determined the degree of differencing (d) which is 2. Now we have to find a degree of the autoregressive and moving average. We can predict the degree of p and q by observing the correlogram plot of the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF). The estimated ARIMA model is (p, d, q) = (1, 2, 2). This model is chosen based on Akaike Information Criterion (AIC) value. ARIMA (1, 2, 2) has the lowest AIC value which is 61687.35. After we find the fit model, the next step is to estimate the parameters. The estimated parameters are given in the table below.

|  |  |  |
| --- | --- | --- |
| ARIMA (1, 2, 2)  AIC: 61687.35 | | |
| Coefficients | Value | Standard error |
| ar1 | 0.6050 | 0.0101 |
| ma1 | -1.8701 | 0.0064 |
| ma2 | 0.8701 | 0.0064 |

**Table 4.1b: Estimated parameters of ARIMA (1, 2, 2) model.**

The last step of fitting an ARIMA model is diagnostic checking. Ljung Box test shows that the model is a white noise with a p-value of 0.8246.

4.1.2 ARIMAX:

ARIMAX model requires exogenous variables. The exogenous variables of our model are ‘Total Cloud Cover’, ‘Dew Point Temperature’, ‘Evaporation’, ‘Solar Radiation’, ‘Mean Sea Level Pressure’, and ‘Wind Speed’. After adding these independent variables ARIMA (1, 2, 2) model is found to be adequate as well. The estimated parameters are given in the table below.

|  |  |  |
| --- | --- | --- |
| ARIMA (1, 2, 2)  AIC = 54672.18 | | |
| Coefficients | Value | Standard Error |
| ar1 | 0.3869 | 0.0114 |
| ma1 | -1.8485 | 0.0076 |
| ma2 | 0.8485 | 0.0076 |
| Total Cloud Cover | -0.292 | 0.018 |
| Dew Point Temperature | 2.4933 | 0.0398 |
| Evaporation | -0.3504 | 0.0161 |
| Solar Radiation | 1.8987 | 0.0707 |
| Mean Sea Level Pressure | -0.1920 | 0.0297 |
| Wind Speed | -0.1027 | 0.0110 |

**Table 4.1c Estimated parameters of ARIMA (1, 2, 2) model with exogenous variables.**

Ljung Box test shows this model is also a white noise with p-value of 0.1306.

### 4.2 Machine Learning Models Diagnostics

#### 4.2.1 Data Multicollinearity

Specific assumptions must be corroborated in the data before machine learning algorithms are applied to a data set. It is crucial to first determine whether there is any relation between the features and the target as well as between the features themselves. The collinearity between the features and the intended data was assessed and displayed in Table 3.1. According to the table, the collinearity between the features was within the tolerable ranges necessary for machine-learning approaches. The collinearity between dew point temperature and mean sea level pressure exceeds the permissible range.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Predictor | Total Cloud Cover | Dew Point Temperature | Evaporation | Mean Sea Level Pressure | Solar Radiation | Wind Speed |
| Total Cloud Cover | 1.00 |  |  |  |  |  |
| Dew Point Temperature | 0.73 | 1.00 |  |  |  |  |
| Evaporation | -0.08 | -0.50 | 1.00 |  |  |  |
| Mean Sea Level Pressure | -0.70 | -0.81 | 0.43 | 1.00 |  |  |
| Solar Radiation | 0.53 | 0.66 | -0.57 | -0.79 | 1.00 |  |
| Wind Speed | 0.46 | 0.30 | 0.12 | -0.40 | 0.27 | 1.00 |

**Table 4.2a: Features collinearity and correlation matrix.**

#### 4.2.2 Mann-Kendall Test

Table 4.2b illustrates the Mann-Kendall (MK) tests that were performed for the time series data. Since the p-value was less than 0.05, the Mann-Kendall (MK) tests' p-values showed trends in temperature as well as in each of the input features over time. The temperature, total cloud cover, dew point temperatures and mean sea level pressure, all of which were already known to have trends, were, therefore, demonstrated to exhibit positive increasing trends due to the positive value of Kendall's τ. On the other hand, evaporation, wind speed, and solar radiation showed declining trends. Sen's Slope, which measures the slope of a trend, revealed a positive trend in the temperature time series of 0.000081 degrees Celsius per day, which is a definite sign of climate change in the study area. Similarly, Sen's Slope was just 0.000073 K each day, which indicated the Dew Point Temperature trend, which had previously been detected as a positive trend. Additionally, the Sen's Slope for the Mean Sea Level Pressure trend, which was similarly shown to be positive, was only 0.003739379 Pa per day.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| M-K test | Sen's Slope | Kendall's τ | p-value | alpha (ɑ) | Test interpretation |
| Temperature | 0.000081 | 0.0903 | <0.01 | 0.05 | Trend in Series |
| Total Cloud Cover | 0 | 0.000679 | 0.88333 | 0.05 | Trend in Series |
| Dew Point Temperature | 7.26853e-05 | 0.093 | <0.01 | 0.05 | Trend in Series |
| Evaporation | -2.872019e-10 | -0.0248 | <0.01 | 0.05 | Trend in Series |
| Solar Radiation | -3.561894 | -0.133 | <0.01 | 0.05 | Trend in Series |
| Mean Sea Level Pressure | 0.003739379 | 0.0268 | <0.01 | 0.05 | Trend in Series |
| Wind Speed | -3.075041e-06 | -0.0227 | <0.01 | 0.05 | Trend in Series |

**Table 4.2b: Mann Kendall test statistics.**

### 4.3 Comparison between time series models and Machine Learning Models

Using the offered feature data sets, classical time series models and machine learning models were utilized to predict the temperature trend in the study area. From January 1, 1965, to December 31, 2021, the combined data collection included 20819 records of daily temperature, total cloud cover, dew point temperatures, evaporation, mean sea level pressure, solar radiation, and wind speed. The training set has 18627 records from January 1, 1965, to December 31, 2015. Data from January 1, 2016, to December 31, 2021, were kept for the testing phase that determined the optimum method. To estimate the temperature for the day of concern, we used the data from the previous day, which included daily temperature, total cloud cover, dew point temperatures, evaporation, mean sea level pressure, solar radiation, and wind speed.

Seven distinct indices, including the coefficient of determination (R2), the index of agreement (d), the Nash-Sutcliffe efficiency (NSE), the mean absolute error (M.A.E.), the root mean square error (RMSE), the percent bias (PBAIS), and the RMSE-observations standard deviation ratio (R.S.R.) were used to examine each trained model's performance in predicting temperature for future dates. The outcomes for each approach were examined more in-depth in this segment.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Methods | RSR | NSE | MAE | Index of Agreement, d | PBIAS (%) | RMSE | R2 |
| ARIMA | 0.5798 | -3.8939 | 7.5739 | 0.7049 | 79.2350 | 8.4407 | 0.0040 |
| ARIMAX | 0.1091 | 0.8268 | 1.2969 | 0.9592 | 0.4844 | 1.5879 | 0.8555 |
| KNN | 0.1667 | 0.5953 | 1.9380 | 0.8521 | 5.5605 | 2.4272 | 0.6258 |
| SVM | 0.1860 | 0.4961 | 2.1586 | 0.8113 | 3.7220 | 2.7085 | 0.5156 |
| Random Forest | 0.0908 | 0.8798 | 1.0506 | 0.9653 | 22.1884 | 1.3226 | 0.9126 |
| Adaptive Boosting | 0.1131 | 0.8138 | 1.3254 | 0.9341 | 5.7312 | 1.6464 | 0.8792 |
| ANN (ReLU) | 0.0771 | 0.9135 | 0.8871 | 0.9742 | 0.6770 | 1.1223 | 0.9275 |

**Table 4.3: Performance ratings of the models for the test dataset.**

The first time series model ARIMA was used to predict the temperature data. It explicitly caters to a suite of standard structures in time series data, and as such provides a simple yet powerful method for making skillful time series forecasts **[Table 4.3].** The relationship between the observed value and the ARIMA models test outputs is shown in **Figure 4.3a**. The plot seems to be less fitted.

A picture containing diagram

Description automatically generated

**Figure 4.3a: Relationship between observed and test outputs for ARIMA**

The ARIMAX model was then used to predict temperature. It may be considered a multiple regression model with one or more autoregressive (AR) components and/or one or more moving average (MA) terms [**Table 4.3**]. The relationship between the observed value and the ARIMAX models test outputs is shown in **Figure 4.3b**. The plot seems to better fit with higher accuracy compared to the ARIMA model.

A picture containing chart

Description automatically generated

**Figure 4.3b: Relationship between observed and test outputs for ARIMAX**

The first machine learning technique KNN was used to predict the temperature data. After several trials of the technique, 23 neighbors (k) were determined to be the appropriate number. The KNN method's chosen distance metric was Euclidian, and uniform weighting was also applied [**Table 4.3**]. The relationship between the observed value and the KNN method's test results is depicted in **Figure 4.3c**. The plot seems to be less fitted compared to the ARIMAX model.

Chart, scatter chart

Description automatically generated

**Figure 4.3c: Relationship between observed and test outputs for KNN**

The SVM method's features were evaluated and tuned by trial and error. For this investigation, the software was allowed to determine the gamma value automatically while the radial basis function was used as the kernel function. [**Table 4.3**]. The relationship between the observed temperature and the SVM method's test results is depicted in **Figure 4.3d**. The plot seems to be less fitted compared to the KNN and ARIMAX methods.

Chart, scatter chart

Description automatically generated

**Figure 4.3d: Relationship between observed and test outputs for SVM**

The Random Forest model served as the study's first ensemble model. 500 trees were chosen as the final number of trees after extensive trial and error [**Table 4.3**]. The relationship between the random forest findings and the observed temperature is seen in **Figure 4.3e**. The plot demonstrates that the model fitted well compared to the previous four methods, and the results were highly accurate.

Chart, scatter chart

Description automatically generated

**Figure 4.3e: Relationship between observed and test outputs for Random Forest**

Adaptive Boosting (AdB) was the second ensemble model that was evaluated. SAMME (Stagewise Additive Modeling utilizing a Multi-class Exponential loss function) was chosen as the classification approach, and a square regression loss function was employed [**Table 4.3**]. The relationship between the AdB findings and the observed temperature is represented in **Figure 3.3e**. The model fitted well and had similar results compared to the random forest model despite having less accuracy.

Chart, scatter chart

Description automatically generated

**Figure 4.3f: Relationship between observed and test outputs for AdB.**

The activation function in a neural network is in function transforming the node's summed weighted input into the activation of the node or output for that input. In our study, a piecewise linear function called the rectified linear activation function (ReLU) was employed [**Table 4.3**]. **Figure 4.3f** illustrates the relationship between the ANN findings and the observed temperature. The model was precise and had the best fit compared to all the time series and machine learning models.

Chart, scatter chart

Description automatically generated

**Figure 4.3g: Relationship between observed and test outputs for ANN**

**Table 4.3** and the plots in **Figure 4.3g** demonstrates that all the procedures were reliable and acceptable except the ARIMA. the findings of the ARIMA model were much worse than those of the other techniques. Based on the R2 value, the best and poorest models were the ANN approach (R2 =0.9275) and the ARIMA method (R2 = 0.0040).

R2 is typically insufficient to compare the results of different machine learning models. As a result, six additional performance indices were employed to contrast the different approaches. The values for the RSR, NSE, MAE, Index of agreement(d), PBIAS, and RSME for each approach are shown in **Table 4.3**. Except for ARIMA and SVM, all the approaches demonstrated acceptable performance, according to (D. N. Moriasi et al., 2007). Based on three measurements (NSE = 0.9135, Index of Agreement(d) = 0.9742, and R2 = 0.9275) ANN Model showed the best performance and the ARIMA model showed the worst performance (NSE=-3.8939, Index of Agreement, d=0.7049 and R2 0.0040) among the techniques. Similarly, according to the second set of three indices the ANN approach (RSR =0.0771, MAE = 0.8871, and RMSE = 1.1223) was the best and ARIMA was the worst (RSR=0.5798, MAE=7.5739, and RMSE=8.4407). Based only on the PBIAS indicator (PBIAS (%) = 0.6770), the ANN model performed best and ARIMA was the worst model (PBIAS(%)=79.2350). For all seven performance measures, the ARIMA approach produced the least reliable findings. Therefore, according to the seven indices ANN models produced the best results in daily temperature prediction compared to all the time series and machine learning models. Similar findings have been reported in other research that compared machine learning models for predicting temperature (Azari et al., 2022).

# CHAPTER V

# 5 CONCLUSIONS AND RECOMMENDATIONS

## 5.1 Introduction

This chapter includes two main sections. In the first section, the main results that have been obtained from the study are presented. The second section summarizes a viable recommendation about global warming.

## 5.2 Findings of the Research

The main findings of the research can be summarized as follows:

The Sylhet region's meteorological time series data, containing temperature, total cloud cover, dew point temperature, evaporation, solar radiation, mean sea level pressure, and wind speed, were extracted for this study. Data were collected between January 1, 1965, and December 31, 2021, or approximately 57 years. The purpose of this study was to assess the capacity of several machine learning (ML) techniques to forecast temperature in order to investigate climate change, as well as to compare them to classical time series models ARIMA, ARIMAX, etc.

### 5.2.1 Time Series Models

Two traditional time series approaches, ARIMA and ARIMAX, were employed in this study. The ARIMAX model ( R2 =0.8555) is well-fitted with higher accuracy compared to the ARIMA model ( R2 =0.0004). Seven indices show that ARIMAX outperforms the ARIMAX models in terms of overall performance for time series analysis (RSR, NSE, MAE,d, PAB, RMSE, RSR, and R2).

### 5.2.2 Machine learning Methods

In order to predict the temperature, six machine learning techniques including kNN, SVM, Random Forest, AdB, and ANN were employed. Data from January 1, 1965, to December 31, 2015, were used as training data.  Then data from January 1, 2016, through December 31, 2021, were used for testing. Except for SVM, all the methods were well-fitted, according to the test findings. The R2 score for SVM was the lowest, coming in at 0.5156. The following six additional performance indicators were used to determine the best-fitted method: RSR, NSE, MAE, Index of Agreement (d), PBIAS (%), and RMSE. However, seven performance indicators showed that ANN was the best method compared to others for predicting temperature.

## 5.3 Comparison between Statistical Methods and Machine Learning Methods

Even though the ARIMAX model outperformed the ARIMA model using seven different indices (RSR, NSE, MAE,d, PAB, RMSE, RSR, and R2). Furthermore, according to seven indices RSR, NSE, MAE,d, PAB, RMSE, RSR, and R2) ARIMAX model outperformed both kNN and SVM machine learning approaches. Using the daily temperature, total cloud cover, dew point temperatures, evaporation, mean sea level pressure, solar radiation, and wind speed as features, it was shown that the ANN machine learning approach was the most accurate way for predicting temperature comparing six machine learning models and two classical time series methods.

## 5.4 Limitations and Recommendations for Future Works

### 5.4.1 Limitations of the Study

1. In this work, models have not been used to forecast the future in real time.
2. Although feature engineering methods have not been used in this research, feature selection is taken into consideration from the literature.
3. Only the Sylhet region was taken into consideration for our study because of budgetary constraints.
4. Satellite data was used because there was a lack of sufficient data in BMD.

### 5.4.2 Recommendations

1. Future climatology studies in the Sylhet region ought to think about using ANN.
2. Future research should incorporate feature engineering methods.
3. Other ways to improve and enhance this study include considering alternative ensemble methods instead of ours.
4. It is necessary to conduct more studies on temperature prediction for vast areas using a similar methodology.

# REFERENCES

Abdel-Aal, R. E. (2004). Hourly temperature forecasting using abductive networks. Engineering Applications of Artificial Intelligence, 17(5), 543–556. https://doi.org/10.1016/j.engappai.2004.04.002

Ahmadi, M. M., Shahriari, H., & Samimi, Y. (2022). A novel robust control chart for monitoring multiple linear profiles in phase II. Communications in Statistics - Simulation and Computation, 51(11), 6257–6268. https://doi.org/10.1080/03610918.2020.1799228

Akter, M., Uddin, M. S., & Haque, A. (2009). Diagnosis and Management of Diabetes Mellitus through a Knowledge-Based System. In C. T. Lim & J. C. H. Goh (Eds.), 13th International Conference on Biomedical Engineering (pp. 1000–1003). Springer. https://doi.org/10.1007/978-3-540-92841-6\_247

Al-Amin, A., Kari, F., & Alam, G. M. (2013). Global warming and climate change: Prospects and challenges toward long-term policies in Bangladesh. Int. J. of Global Warming, 5, 67–83. https://doi.org/10.1504/IJGW.2013.051483

Andrade, J. R., & Bessa, R. J. (2017). Improving Renewable Energy Forecasting With a Grid of Numerical Weather Predictions. IEEE Transactions on Sustainable Energy, 8(4), 1571–1580. https://doi.org/10.1109/TSTE.2017.2694340

Anjali, T., Chandini, K., Anoop, K., & Lajish, V. L. (2019). Temperature Prediction using Machine Learning Approaches. 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), 1, 1264–1268. https://doi.org/10.1109/ICICICT46008.2019.8993316

Ardabili, S., Mosavi, A., & Várkonyi-Kóczy, A. R. (2020). Advances in Machine Learning Modeling Reviewing Hybrid and Ensemble Methods. In A. R. Várkonyi-Kóczy (Ed.), Engineering for Sustainable Future (pp. 215–227). Springer International Publishing. https://doi.org/10.1007/978-3-030-36841-8\_21

Awal, M., & Khan, M. (2020). Global Warming and Sea Level Rising: Impact on Agriculture and Food Security in Southern Coastal Region of Bangladesh. Asian Journal of Geographical Research, 3, 9–36. https://doi.org/10.9734/AJGR/2020/v3i330107

Azari, B., Hasan, K., Pierce, J., & Ebrahimi, S. (2022). Evaluation of Machine Learning Methods Application in Temperature Prediction. COMPUTATIONAL RESEARCH PROGRESS IN APPLIED SCIENCE & ENGINEERING, 8, 1–12. https://doi.org/10.52547/crpase.8.1.2747

Bang, S., Bishnoi, R., Chauhan, A. S., Dixit, A. K., & Chawla, I. (2019). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. 2019 Twelfth International Conference on Contemporary Computing (IC3), 1–6. https://doi.org/10.1109/IC3.2019.8844901

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, L., & Lai, X. (2011). Comparison between ARIMA and ANN Models Used in Short-Term Wind Speed Forecasting. 2011 Asia-Pacific Power and Energy Engineering Conference, 1–4. https://doi.org/10.1109/APPEEC.2011.5748446

Chou, J.-S., & Tran, D.-S. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy, 165, 709–726. https://doi.org/10.1016/j.energy.2018.09.144

Choudhury, A. M., Haque, M. A., & Quadir, D. A. (1997). Consequences of global warming and sea level rise in Bangladesh. Marine Geodesy, 20(1), 13–31. https://doi.org/10.1080/01490419709388092

Cifuentes, J., Marulanda, G., Bello, A., & Reneses, J. (2020). Air Temperature Forecasting Using Machine Learning Techniques: A Review. Energies, 13(16), Article 16. https://doi.org/10.3390/en13164215

Colak, I., Yesilbudak, M., Genc, N., & Bayindir, R. (2015). Multi-period Prediction of Solar Radiation Using ARMA and ARIMA Models. 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), 1045–1049. https://doi.org/10.1109/ICMLA.2015.33

Cox, D. T. C., Maclean, I. M. D., Gardner, A. S., & Gaston, K. J. (2020). Global variation in diurnal asymmetry in temperature, cloud cover, specific humidity and precipitation and its association with leaf area index. Global Change Biology, 26(12), 7099–7111. https://doi.org/10.1111/gcb.15336

D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, & T. L. Veith. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE, 50(3), 885–900. https://doi.org/10.13031/2013.23153

Dimri, T., Ahmad, S., & Sharif, M. (2020). Time series analysis of climate variables using seasonal ARIMA approach. Journal of Earth System Science, 129(1), 149. https://doi.org/10.1007/s12040-020-01408-x

Duan, S., Yang, W., Wang, X., Mao, S., & Zhang, Y. (2019). Forecasting of Grain Pile Temperature From Meteorological Factors Using Machine Learning. IEEE Access, 7, 130721–130733. https://doi.org/10.1109/ACCESS.2019.2940266

Fotovvati, B., Balasubramanian, M., & Asadi, E. (2020). Modeling and Optimization Approaches of Laser-Based Powder-Bed Fusion Process for Ti-6Al-4V Alloy. Coatings, 10(11), Article 11. https://doi.org/10.3390/coatings10111104

Hastie, T., Friedman, J., & Tibshirani, R. (2001). The Elements of Statistical Learning. Springer. https://doi.org/10.1007/978-0-387-21606-5

Helfer, F., Lemckert, C., & Zhang, H. (2012). Impacts of climate change on temperature and evaporation from a large reservoir in Australia. Journal of Hydrology, 475, 365–378. https://doi.org/10.1016/j.jhydrol.2012.10.008

Huntington, T. G. (2006). Evidence for intensification of the global water cycle: Review and synthesis. Journal of Hydrology, 319(1), 83–95. https://doi.org/10.1016/j.jhydrol.2005.07.003

Islam, Md. N., & Uyeda, H. (2007). Use of TRMM in determining the climatic characteristics of rainfall over Bangladesh. Remote Sensing of Environment, 108(3), 264–276. https://doi.org/10.1016/j.rse.2006.11.011

Khan, M. M. H., Bryceson, I., Kolivras, K. N., Faruque, F., Rahman, M. M., & Haque, U. (2015). Natural disasters and land-use/land-cover change in the southwest coastal areas of Bangladesh. Regional Environmental Change, 15(2), 241–250. https://doi.org/10.1007/s10113-014-0642-8

Lee, J. Y., & Kim, S. (2019). Forecasting daily peak load by time series model with temperature and special days effect. The Korean Journal of Applied Statistics, 32(1), 161–171. https://doi.org/10.5351/KJAS.2019.32.1.161

Li, S., Goel, L., & Wang, P. (2016). An ensemble approach for short-term load forecasting by extreme learning machine. Applied Energy, 170, 22–29. https://doi.org/10.1016/j.apenergy.2016.02.114

Liu, S., Tai, H., Ding, Q., Li, D., Xu, L., & Wei, Y. (2013). A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. Mathematical and Computer Modelling, 58(3), 458–465. https://doi.org/10.1016/j.mcm.2011.11.021

Mahabub, A., Innat, M., & Faruque, M. (2019). A robust system for message filtering using an ensemble machine learning supervised approach. ICIC Express Letters, 10, 805–811. https://doi.org/10.24507/icicelb.10.09.805

Marill, K. A. (2004). Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression. Academic Emergency Medicine, 11(1), 94–102. https://doi.org/10.1197/j.aem.2003.09.006

Mehri, S., Ahmadi, M. M., Shahriari, H., & Aghaie, A. (2021). Robust process capability indices for multiple linear profiles. Quality and Reliability Engineering International, 37(8), 3568–3579. https://doi.org/10.1002/qre.2934

Mohamed, Z. E. (2019). Using the artificial neural networks for prediction and validating solar radiation. Journal of the Egyptian Mathematical Society, 27(1), 47. https://doi.org/10.1186/s42787-019-0043-8

Mortuza, Md. R., Selmi, S., Khudri, Md. M., Ankur, A. K., & Rahman, Md. M. (2014). Evaluation of temporal and spatial trends in relative humidity and dew point temperature in Bangladesh. Arabian Journal of Geosciences, 7(12), 5037–5050. https://doi.org/10.1007/s12517-013-1139-3

Mousavi, S. S., Firouzmand, M., Charmi, M., Hemmati, M., Moghadam, M., & Ghorbani, Y. (2019). Blood pressure estimation from appropriate and inappropriate PPG signals using A whole-based method. Biomedical Signal Processing and Control, 47, 196–206. https://doi.org/10.1016/j.bspc.2018.08.022

Najah, A., El-Shafie, A., Karim, O. A., & El-Shafie, A. H. (2013). Application of artificial neural networks for water quality prediction. Neural Computing and Applications, 22(1), 187–201. https://doi.org/10.1007/s00521-012-0940-3

Nury, A., Alam, Dr. Md. J., & Koch, M. (2013). Time series analysis and Forecasting of Temperatures in the Sylhet Division of Bangladesh.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P., Dubash, N. K., Edenhofer, O., Elgizouli, I., Field, C. B., Forster, P., Friedlingstein, P., Fuglestvedt, J., Gomez-Echeverri, L., Hallegatte, S., … van Ypserle, J.-P. (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In R. K. Pachauri & L. Meyer (Eds.), EPIC3Geneva, Switzerland, IPCC, 151 p., pp. 151, ISBN: 978-92-9169-143-2 (p. 151). IPCC. https://epic.awi.de/id/eprint/37530/

Pachauri, R. K., & Reisinger, A. (2008). Climate change 2007. Synthesis report. Contribution of Working Groups I, II and III to the fourth assessment report. https://www.osti.gov/etdeweb/biblio/944235

Pham, L. T., Luo, L., & Finley, A. (2021). Evaluation of random forests for short-term daily streamflow forecasting in rainfall- and snowmelt-driven watersheds. Hydrology and Earth System Sciences, 25(6), 2997–3015. https://doi.org/10.5194/hess-25-2997-2021

Pour, S. H., Shahid, S., Chung, E.-S., & Wang, X.-J. (2018). Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh. Atmospheric Research, 213, 149–162. https://doi.org/10.1016/j.atmosres.2018.06.006

Rahman, M. A., Yunsheng, L., & Sultana, N. (2017). Analysis and prediction of rainfall trends over Bangladesh using Mann–Kendall, Spearman’s rho tests and ARIMA model. Meteorology and Atmospheric Physics, 129(4), 409–424. https://doi.org/10.1007/s00703-016-0479-4

Räisänen, J., & Alexandersson, H. (2003). A probabilistic view on recent and near future climate change in Sweden. Tellus A: Dynamic Meteorology and Oceanography, 55(2), 113–125. https://doi.org/10.3402/tellusa.v55i2.12089

Rajeevan, M., Pai, D. S., Anil Kumar, R., & Lal, B. (2007). New statistical models for long-range forecasting of southwest monsoon rainfall over India. Climate Dynamics, 28(7), 813–828. https://doi.org/10.1007/s00382-006-0197-6

Rasamoelina, A. D., Adjailia, F., & Sinčák, P. (2020). A Review of Activation Function for Artificial Neural Network. 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI), 281–286. https://doi.org/10.1109/SAMI48414.2020.9108717

Rina, K., Datta, P. S., Singh, C. K., & Mukherjee, S. (2012). Characterization and evaluation of processes governing the groundwater quality in parts of the Sabarmati basin, Gujarat using hydrochemistry integrated with GIS. Hydrological Processes, 26(10), 1538–1551. https://doi.org/10.1002/hyp.8284

Sawan, Z. M. (2018). Climatic variables: Evaporation, sunshine, relative humidity, soil and air temperature and its adverse effects on cotton production. Information Processing in Agriculture, 5(1), 134–148. https://doi.org/10.1016/j.inpa.2017.09.006

Schaller, N., Kay, A. L., Lamb, R., Massey, N. R., van Oldenborgh, G. J., Otto, F. E. L., Sparrow, S. N., Vautard, R., Yiou, P., Ashpole, I., Bowery, A., Crooks, S. M., Haustein, K., Huntingford, C., Ingram, W. J., Jones, R. G., Legg, T., Miller, J., Skeggs, J., … Allen, M. R. (2016). Human influence on climate in the 2014 southern England winter floods and their impacts. Nature Climate Change, 6(6), Article 6. https://doi.org/10.1038/nclimate2927

Shafin, A. A. (2019). Machine Learning Approach to Forecast Average Weather Temperature of Bangladesh. Global Journal of Computer Science and Technology, 19(D3), 39–48.

Shakti, S. P., Hassan, M. K., Zhenning, Y., Caytiles, R. D., & N.Ch.S.N, I. (2017). Annual Automobile Sales Prediction Using ARIMA Model. International Journal of Hybrid Information Technology, 10(6), 13–22. https://doi.org/10.14257/ijhit.2017.10.6.02

Singh, K. P., Basant, A., Malik, A., & Jain, G. (2009). Artificial neural network modeling of the river water quality—A case study. Ecological Modelling, 220(6), 888–895. https://doi.org/10.1016/j.ecolmodel.2009.01.004

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., & Murphy, J. M. (2007). Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model. Science, 317(5839), 796–799. https://doi.org/10.1126/science.1139540

Stammer, D., & Hüttemann, S. (2008). Response of Regional Sea Level to Atmospheric Pressure Loading in a Climate Change Scenario. Journal of Climate, 21(10), 2093–2101. https://doi.org/10.1175/2007JCLI1803.1

Time Series Analysis: Forecasting and Control - ProQuest. (n.d.). Retrieved December 31, 2022, from https://www.proquest.com/openview/3fffd089bf646234e73de5610626ebd5/1?pq-origsite=gscholar&cbl=40685

Tol, R. S. J. (2002). Estimates of the Damage Costs of Climate Change. Part 1: Benchmark Estimates. Environmental and Resource Economics, 21(1), 47–73. https://doi.org/10.1023/A:1014500930521

Tol, R. S. J. (2013). The economic impact of climate change in the 20th and 21st centuries. Climatic Change, 117(4), 795–808. https://doi.org/10.1007/s10584-012-0613-3

Wang, X., Ma, L., & Wang, X. (2010). Apply semi-supervised support vector regression for remote sensing water quality retrieving. 2010 IEEE International Geoscience and Remote Sensing Symposium, 2757–2760. https://doi.org/10.1109/IGARSS.2010.5653832

Yang, H., Chan, L., & King, I. (2002). Support Vector Machine Regression for Volatile Stock Market Prediction. In H. Yin, N. Allinson, R. Freeman, J. Keane, & S. Hubbard (Eds.), Intelligent Data Engineering and Automated Learning—IDEAL 2002 (pp. 391–396). Springer. https://doi.org/10.1007/3-540-45675-9\_58

Zaman, Y. (2018). Machine Learning Model on Rainfall—A Predicted Approach for Bangladesh [Thesis, United International University]. http://dspace.uiu.ac.bd/handle/52243/178